



Aura Science Meeting Data Systems Working Group HIRDLS SIPS Sept. 13, 2006

**Vince Dean, Brendan Torpy, Greg Young
Univ. of Colorado, Boulder**

**Cheryl Craig
NCAR**

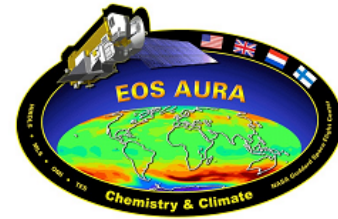
Overview



- **We use big iron for science processing – large, multiprocessor machines.**
- **We use a small PC cluster for SIPS data management and user interface.**
- **We share resources between development and production (SCF and SIPS).**
- **We support experimental runs in production environment, carefully tracked.**
- **Some versions have been released to DAAC and AVDC, others in internal review.**
- **Proliferation of processor versions leads to considerable manual book-keeping and stresses the system design.**
- **DAAC evolution: technical interfaces work, still leaving concern about changes in semantics and metadata.**



Big Iron for processing– shared by production and development



Name	Use	Vendor	CPUs	Memory (Gbytes)	Architecture	OS
hir1 <being retired>	Development and Production	SGI	32	52	MIPS	Irix
hal	Development	SGI	12	24	Itanium	SUSE Linux
hcl	Mostly Production	SGI	80	160	Itanium	SUSE Linux
<new>	Development and backup Production	SGI	32	64	Itanium	SUSE Linux



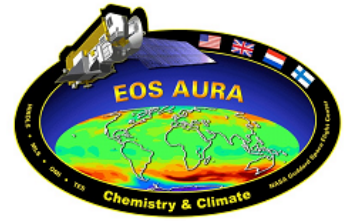
PC Cluster for SIPS Data Management



- **Five commodity PCs**
- **Fedora Linux**
- **Java**
- **Open source tools**
 - **Ant**
 - **JBoss**
 - **MySQL**
 - **Struts**
 - **...**



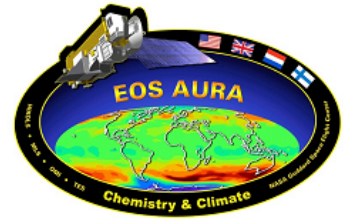
Processing Rate



- **Process one day, end-to-end, in 9 hours.**
- **Effective throughput, running parallel jobs:**
 - One day processed every 2 hours
 - 12x processing
 - 2 years re-processed in ~2 months



Goddard DAAC Evolution



- **DAAC is generally honoring existing interfaces for distributions and has been responsive to our requests.**
- **Passed mini-MOSS test, ingesting GEOS-5 data from S4PA.**
- **Lack of “database ID” in distribution notice has required some reprogramming.**
- **Are there other changes in semantics waiting for us??**
- **Metadata**
 - **Fewer explicit requirements.**
 - **Reevaluate scope of metadata—what will users need?**



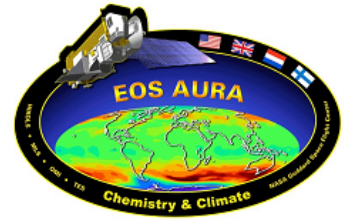
Experimentation in SIPS



- **HIRDLS obstruction required a new round of highly experimental development.**
- **We have chosen to install many experimental processors in our SIPS system and have run many one-off tests.**
- **The automation and audit trail have been indispensable for that work.**
- **QA plots and comparisons with other products are generated automatically.**
- **We have been able to handle the large number of experimental versions, but it requires considerable manual book-keeping.**
- **We plan enhancements to simplify data management.**



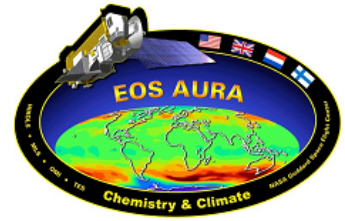
Data and Processor Versioning



- **Versions of our data products are distinguished by:**
 - **Association with processor versions that created them**
 - **Naming conventions:**
 - **2.00, 2.01, etc.**
 - Suitable for release to outside users--GES DISC, and/or AVDC.
 - **2.02.01, 2.02.02, etc.**
 - Internal, development versions.
- **Versions of external products are harder to track. Strategies include:**
 - **Newest is best—ignore old versions.**
 - Generally suitable for level 0 data, attitude and ephemeris.
 - **New file type for each new version of external data.**
 - We are using this for MLS products, to distinguish v1.5 from v2.1.
- **Work in progress.**



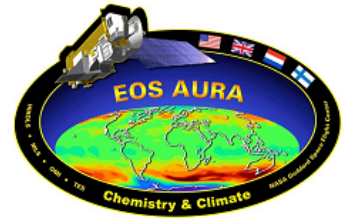
Scientists as SIPS users



- **SIPS was designed to ingest files, run production jobs and deliver results to the DAAC, with the operator as the primary user.**
- **Science users find SIPS useful to:**
 - Track experimental jobs
 - View QA plots
- **Generally well received, in spite of complex user interface.**



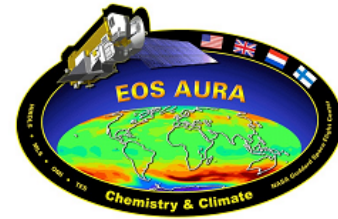
Agile Development



- **Two developers maintaining 100,000+ lines of code, with agile practices, including:**
 - **Frequent releases**
 - **Extensive unit tests**
- **We are able to respond to:**
 - **need for experimental runs**
 - **changes for DAAC evolution**



Data Releases



- **Version 2.00**
 - Delivered 27 days of L2 data to Goddard DISC and AVDC.
- **Version 2.01**
 - Delivered additional 10 days L2 data to AVDC.
- **Processing remains experimental, for selected day of interest from each new software version.**
- **We do not yet have plans for wholesale reprocessing.**
- **HIRDLS Documents at Goddard DISC on Aura documentation page:**
 - A short guide to the use and interpretation of V2.00 Level 2 data.
 - Data Description and Quality -- Version 2.00
- **<http://daac.gsfc.nasa.gov/Aura/documentation/index.shtml>**



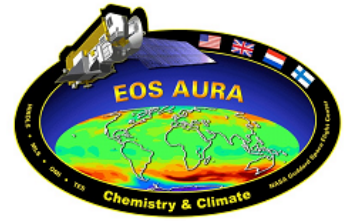
Scan Tables



- **HIRDLS instrument has used four different scan patterns (scan tables) since January 2005, each one designed to allow us to compensate better for the obstruction.**
 - **ST 30 – January 21 ... April 28, 2005**
 - **ST 13 – April 28, 2005 ... April 24, 2006**
 - **ST 22 – April 24 ... May 4, 2006**
 - **ST 23 – May 4, 2006 ...**
- **Recent versions of the de-oscillation code have custom features for each scan table. Each of several recent releases has added the ability to handle one more scan table.**



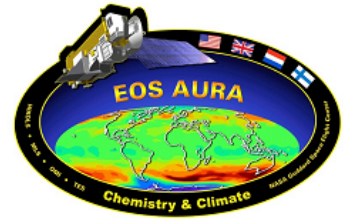
Version 2.00



- **Installed in June, 2006**
- **Recent versions of de-oscillation code are customized for each scan table.**
- **Handles only scan table 23.**
 - **ST 23 – May 4, 2006 ...**
- **Delivered 27 days L2 data to AVDC and Goddard DAAC**
 - **All scan table 23**
 - **May 4 .. 31, 2006**
 - **except May 23 – pitch up**



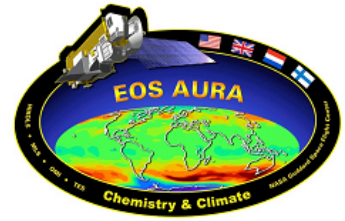
Version 2.01



- **Installed in July, 2006**
- **Adds ability to process scan table 22**
 - **ST 22 – April 24 ... May 4, 2006**
 - **ST 23 – May 4, 2006 ...**
- **Processed and delivered 10 additional days L2 data to AVDC**
 - **All scan table 22**
 - **April 25 ... May 4, 2006**



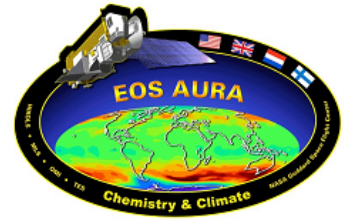
Version 2.02



- **Installed in August, 2006**
- **Adds improvements to cloud detection algorithms**
- **Handles scan tables 22 and 23**
 - **ST 22 – April 24 ... May 4, 2006**
 - **ST 23 – May 4, 2006 ...**
- **Processed 47 selected days of interest from**
 - **Scan tables 22 and 23**
 - **Spanning April 25 ... August 12, 2006**



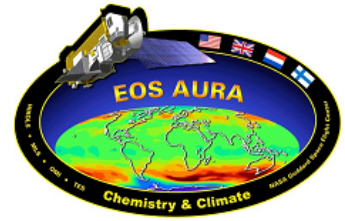
Version 2.02.02



- **Installed in September, 2006**
- **Adds support for scan table 13**
 - **ST 13 – April 28, 2005 ... April 24, 2006**
 - **ST 22 – April 24 ... May 4, 2006**
 - **ST 23 – May 4, 2006 ...**
- **Processed 27 additional days of interest**
 - **Scan table 13.**
 - **Spanning May 5, 2005 ... April 30, 2006**



Version 2.02.03



- **Installed in September, 2006**
- **Adds support for scan table 30**
 - **ST 30 – January 21 ... April 28, 2005**
 - **ST 13 – April 28, 2005 ... April 24, 2006**
 - **ST 22 – April 24 ... May 4, 2006**
 - **ST 23 – May 4, 2006 ...**
- **Processed 12 additional days of interest to coincide with PAVE mission.**
 - **Scan table 30**
 - **May 5, 2005 ... April 30, 2006**